

Multidirectional Subspace Expansion for One-Parameter and Multiparameter Tikhonov Regularization

Ian N. Zwaan¹ · Michiel E. Hochstenbach¹

Received: 10 December 2015 / Revised: 11 August 2016 / Accepted: 13 August 2016 /
Published online: 24 August 2016
© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract Tikhonov regularization is a popular method to approximate solutions of linear discrete ill-posed problems when the observed or measured data is contaminated by noise. Multiparameter Tikhonov regularization may improve the quality of the computed approximate solutions. We propose a new iterative method for large-scale multiparameter Tikhonov regularization with general regularization operators based on a multidirectional subspace expansion. The multidirectional subspace expansion may be combined with subspace truncation to avoid excessive growth of the search space. Furthermore, we introduce a simple and effective parameter selection strategy based on the discrepancy principle and related to perturbation results.

Keywords Tikhonov · Multiparameter Tikhonov · Generalized Krylov · Multidirectional subspace expansion · Subspace truncation · Subspace method · Linear discrete ill-posed problem · Regularization · Regularization parameter

Mathematics Subject Classification 15A29 · 65F10 · 65F22 · 65F30 · 65R30 · 65R32

This work is part of the research program “Innovative methods for large matrix problems”, which is (partly) financed by the Netherlands Organization for Scientific Research (NWO).

✉ Ian N. Zwaan
i.n.zwaan@tue.nl

Michiel E. Hochstenbach
<http://www.win.tue.nl/~hochsten>

¹ Department of Mathematics and Computer Science, Eindhoven University of Technology,
PO Box 513, 5600 MB Eindhoven, The Netherlands

1 Introduction

We consider one-parameter and multiparameter Tikhonov regularization problems of the form

$$\operatorname{argmin}_x \|Ax - b\|^2 + \sum_{i=1}^{\ell} \mu^i \|L^i x\|^2 \quad (\ell \geq 1), \quad (1)$$

where $\|\cdot\|$ denotes the 2-norm and the superscript i is used as an index. We focus on large-scale discrete ill-posed problems such as the discretization of Fredholm integral equations of the first kind. More precisely, assume A is an ill-conditioned or even singular $m \times n$ matrix with $m \geq n$, L^i are $p^i \times n$ matrices such that the nullspaces of A and L^i intersect trivially, and μ^i are nonnegative regularization parameters. Furthermore, assume b is contaminated by an error e and satisfies $b = Ax_* + e$, where x_* is the exact solution. Finally, we assume that a bound $\|e\| \leq \varepsilon$ is available, so that the discrepancy principle can be used.

In one-parameter Tikhonov regularization ($\ell = 1$), the choice of the regularization operator is typically significant, since frequencies in the nullspace of the operator remain unpenalized. Multiparameter Tikhonov can be used when a satisfactory choice of the regularization operator is unknown in advance, or can be seen as an attempt to combine the strengths of different regularization operators. In some applications, using more than one regularization operator and parameter allows for more accurate solutions [1, 2, 17, 20].

Solving (1) for large-scale problems may be challenging. In case the μ^i are fixed *a priori*, methods such as LSQR [21] or LSMR [4] may be used. However, the problem becomes more complicated when the regularization parameters are not fixed in advance [12, 15, 17]. In this paper, we present a new subspace method consisting of three phases; a new expansion phase, a new extraction phase, and a new truncation phase. To be more specific, let $\mathcal{X}_k \subset \mathbb{R}^n$ be a subspace of dimension $k \ll n$, and let the columns of X_k form an orthonormal basis for \mathcal{X}_k . Then we can compute matrix decompositions

$$\begin{aligned} AX_k &= U_{k+1} \underline{H}_k \\ L^i X_k &= V_k^i K_k^i \quad (i = 1, 2, \dots, \ell), \end{aligned} \quad (2)$$

where U_{k+1} and V_k^i are have orthonormal columns, $\beta u_1 = b$, $\beta = \|b\|$, \underline{H}_k is a $(k+1) \times k$ Hessenberg matrix, and K_k^i is upper triangular. Denote $\mu = (\mu^1, \dots, \mu^\ell)$ for convenience. Now restrict the solution space to \mathcal{X}_k so that $x_k(\mu) = X_k c_k(\mu)$, where

$$\begin{aligned} c_k(\mu) &= \operatorname{argmin}_c \|AX_k c - b\|^2 + \sum_{i=1}^{\ell} \mu^i \|L^i X_k c\|^2 \\ &= \operatorname{argmin}_c \|\underline{H}_k c - \beta e_1\|^2 + \sum_{i=1}^{\ell} \mu^i \|K_k^i c\|^2. \end{aligned} \quad (3)$$

The vector e_1 is the first standard basis vector of appropriate dimension. Our paper has three contributions. First, a new expansion phase where we add multiple search directions to \mathcal{X}_k . Second, a new truncation phase which removes unwanted new search directions. Third, a new method for selecting the regularization parameters μ_k^i in the extraction phase. The three phases work alongside each other: the intermediate solution obtained in the extraction phase is preserved in the truncation phase, whereas the remaining perpendicular component(s) from the expansion phase are removed.

The paper is organized as follows. In Sect. 2 an existing nonlinear subspace method is discussed, whereafter we propose the new multidirectional subspace expansion of the expansion phase. Discussion of the truncation phase follows immediately. Section 3 is focused on discrepancy principle based parameter selection for one-parameter regularization. New lower and upper bounds on the regularization parameter are provided. Sections 4 and 5 describe the extraction phase. In the former, a straightforward parameter selection strategy for multiparameter regularization is given, in the latter, a justification using perturbation analysis. Numerical experiments are performed in Sect. 6 and demonstrate the competitiveness of our new method. We end with concluding remarks in Sect. 7.

2 Subspace Expansion for Multiparameter Tikhonov

Let us first consider one-parameter Tikhonov regularization with a general regularization operator. Then $\ell = 1$ and we write $\mu = \mu^1$, $L = L^1$, and $K_k = K_k^1$, such that (1) simplifies to

$$\operatorname{argmin}_x \|Ax - b\|^2 + \mu \|Lx\|^2.$$

When $L = I$ we use the Golub–Kahan–Lanczos bidiagonalization procedure to generate the Krylov subspace

$$\mathcal{K}_k = \mathcal{K}_k(A^*A, A^*b) = \operatorname{span}\{A^*b, (A^*A)A^*b, \dots, (A^*A)^{k-1}A^*b\}.$$

In this case \underline{H}_k is lower bidiagonal and K_k is the identity and

$$x_{k+1} = \frac{(I - X_k X_k^*)A^*u_{k+1}}{\|(I - X_k X_k^*)A^*u_{k+1}\|}$$

If $L \neq I$ one can still try to use the above Krylov subspace [12], however, it may be more natural to consider a shift-independent generalized Krylov subspace of the form

$$\mathcal{K}_k = \mathcal{K}_k(A^*A, L^*L, A^*b),$$

spanned by the first k vectors in

$$\begin{aligned} \text{Group 0} & \quad A^*b \\ \text{Group 1} & \quad (A^*A)A^*b, (L^*L)A^*b \\ \text{Group 2} & \quad (A^*A)^2A^*b, (A^*A)(L^*L)A^*b, (L^*L)(A^*A)A^*b, (L^*L)^2A^*b \\ & \quad \dots \end{aligned}$$

This generalized Krylov subspace was first studied by Li and Ye [18] and later by Reichel et al. [23]. An orthonormal basis can be created with a generalization of Golub–Kahan–Lanczos bidiagonalization [13]. However, while the search space grows linearly as a function of the number of matrix-vector products, the dimension of the generalized Krylov subspace grows exponentially as a function of the total degree of a bivariate matrix polynomial. As a result, if we take any vector $x \in \mathcal{K}_k(A^*A, L^*L, A^*b)$ and write it as $p(A^*A, L^*L)A^*b$, where p is a bivariate polynomial, then p has at most degree $\lceil \log_2 k \rceil$. This low degree may be undesirable especially for small regularization parameters μ . Reichel and Yu [24, 25] solve this in part with algorithms that can prioritize one operator over the other. For instance, if w is a vector in a group j and B has priority over A , then group $j + 1$ contains $(A^*A)w$,

$(B^*B)\mathbf{w}$, $(B^*B)^2\mathbf{w}$, ..., $(B^*B)^\rho\mathbf{w}$. The downside is that ρ is a user defined constant, and that the expansion vectors are not necessarily optimal.

An alternative approach is a greedy nonlinear method described by Lampe et al. [17]. We briefly review their method and state a straightforward extension to multiparameter Tikhonov regularization. First note that the low-dimensional minimization in (3) simplifies to

$$\begin{aligned} \mathbf{c}_k(\mu) &= \underset{\mathbf{c}}{\operatorname{argmin}} \|AX_k\mathbf{c} - \mathbf{b}\|^2 + \mu\|LX_k\mathbf{c}\|^2 \\ &= \underset{\mathbf{c}}{\operatorname{argmin}} \|\underline{H}_k\mathbf{c} - \beta\mathbf{e}_1\|^2 + \mu\|K_k\mathbf{c}\|^2, \end{aligned}$$

in the one-parameter case. Next, compute a value $\mu = \mu_k$ using, e.g., the discrepancy principle. It is easy to verify that

$$\begin{aligned} A^*\mathbf{b} - (A^*A + \mu_k L^*L)\mathbf{x}_k(\mu_k) \\ = A^*U_{k+1}(\beta\mathbf{e}_1 - \underline{H}_k\mathbf{c}_k(\mu_k)) + \mu_k L^*V_k K_k\mathbf{c}_k(\mu_k) \end{aligned}$$

is perpendicular to \mathcal{X}_k , as well as the gradient of the cost function

$$\mathbf{x} \mapsto \frac{1}{2}(\|A\mathbf{x} - \mathbf{b}\|^2 + \mu\|L\mathbf{x}\|^2)$$

in the point $\mathbf{x}_k(\mu_k)$. Therefore, this vector is used to expand the search space. As usual, expansion and extraction are repeated until suitable stopping criteria are met.

As previously stated, Lampe et al. [17] consider only one-parameter Tikhonov regularization, however, their method readily extends to multiparameter Tikhonov regularization. Again, the first step is to decide on regularization parameters μ_k . Next, use the residual of the normal equations

$$\begin{aligned} A^*\mathbf{b} - \left(A^*A + \sum_{i=1}^{\ell} \mu_k^i L^{i*}L^i\right)\mathbf{x}_k(\mu_k) \\ = A^*U_{k+1}(\beta\mathbf{e}_1 - \underline{H}_k\mathbf{c}_k(\mu_k)) - \sum_{i=1}^{\ell} \mu_k^i L^{i*}V_k^i K_k^i\mathbf{c}_k(\mu_k), \end{aligned}$$

to expand the search space. Note that the residual is again orthogonal to \mathcal{X}_k as well as the gradient of the cost function

$$\mathbf{x} \mapsto \frac{1}{2}\left(\|A\mathbf{x} - \mathbf{b}\|^2 + \sum_{i=1}^{\ell} \mu^i \|L^i\mathbf{x}\|^2\right).$$

We summarize this multiparameter method in Algorithm 1, but remark that in practice we initially use Golub–Kahan–Lanczos bidiagonalization until a μ_k can be found that satisfies the discrepancy principle.

Algorithm 1 (Generalized Krylov subspace Tikhonov regularization; extension of [17])

Input: Measurement matrix A , regularization operators L^1, \dots, L^ℓ , and data \mathbf{b} .

Output: Approximate solution $\mathbf{x}_k \approx \mathbf{x}_*$.

1. Initialize $\beta = \|\mathbf{b}\|$, $U_1 = \mathbf{b}/\beta$, $X_0 = []$, $\mathbf{x}_0 = \mathbf{0}$, and $\mu_0 = 0$.
 for $k = 1, 2, \dots$ **do**
2. Expand X_{k-1} with $A^*\mathbf{b} - (A^*A + \sum_{i=1}^{\ell} \mu_{k-1}^i L^{i*}L^i)\mathbf{x}_{k-1}$.

3. Update $AX_k = U_{k+1}\underline{H}_k$ and $L^i X_k = V_k^i K_k^i$.
4. Select μ_k ; see Sect. 4 and Algorithm 3.
5. $\mathbf{c}_k = \underset{\mathbf{c}}{\operatorname{argmin}} \left\| \left[\underline{H}_k; \sqrt{\mu_k^1} K_k^1; \dots; \sqrt{\mu_k^\ell} K_k^\ell \right] \mathbf{c} - \beta \mathbf{e}_1 \right\|$.
6. $\mathbf{x}_k = X_k \mathbf{c}_k$.
7. **if** $\|\mathbf{x}_k - \mathbf{x}_{k-1}\|/\|\mathbf{x}_k\|$ is sufficiently small **then break**

Suitable regularization operators often depend on the problem and its solution. Multiparameter regularization may be used when a priori information is lacking. In this case, it is not obvious that the residual vector above is a “good” expansion vector, in particular if the intermediate regularization parameters μ_k are not necessarily accurate. Hence, we propose to remove the dependence on the parameters to some extent by expanding the search space with the vectors

$$A^* A \mathbf{x}_k(\mu_k), \quad L^{1*} L^1 \mathbf{x}_k(\mu_k), \quad \dots, \quad L^{\ell*} L^\ell \mathbf{x}_k(\mu_k), \quad (4)$$

separately. Here, we omit $A^* \mathbf{b}$ as it is already contained in X_k . Since we expand the search space in multiple directions, we refer to this expansion as a “multidirectional” subspace expansion. Observe that the previous residual expansion vector is in the span of the multidirectional expansion vectors.

It is unappealing for the search space to grow with $\ell + 1$ basis vectors per iteration, because the cost of orthogonalization and the cost of solving the projected problems depend on the dimension of the search space. Therefore, we wish to condense the best portions of the multiple directions in a single vector, and use the following approach. First we expand X_k with the vectors in (4) and obtain $\tilde{X}_{k+\ell+1}$. Then we compute the decompositions

$$\begin{aligned} A \tilde{X}_{k+\ell+1} &= \tilde{U}_{k+\ell+2} \tilde{\underline{H}}_{k+\ell+1} \\ L^i \tilde{X}_{k+\ell+1} &= \tilde{V}_{k+\ell+1}^i \tilde{K}_{k+\ell+1}^i \quad (i = 1, 2, \dots, \ell), \end{aligned}$$

analogously to (2) and determine parameters μ_{k+1} and the approximate solution $\tilde{\mathbf{c}}_{k+\ell+1}$. Next, we compute

$$\begin{aligned} A(\tilde{X}_{k+\ell+1} Z^*) &= (\tilde{U}_{k+\ell+2} P^*)(P \tilde{\underline{H}}_{k+\ell+1} Z^*) \\ L^i(\tilde{X}_{k+\ell+1} Z^*) &= (\tilde{V}_{k+\ell+1}^i Q^i)(Q^i \tilde{K}_{k+\ell+1}^i Z^*) \quad (i = 1, 2, \dots, \ell), \end{aligned} \quad (5)$$

where Z , P , and Q^i orthonormal matrices of the form

$$Z = \begin{bmatrix} I_k & \\ & Z_{\ell+1} \end{bmatrix}, \quad P = \begin{bmatrix} I_{k+1} & \\ & P_{\ell+1} \end{bmatrix}, \quad Q^i = \begin{bmatrix} I_k & \\ & Q_{\ell+1}^i \end{bmatrix}. \quad (6)$$

Here I_k is the $k \times k$ identity matrix and $Z_{\ell+1}$ is an orthonormal matrix so that $Z_{\ell+1} \tilde{\mathbf{c}}_{k+1:k+\ell+1} = \gamma \mathbf{e}_1$ for some scalar γ . The matrices $P_{\ell+1}$ and $Q_{\ell+1}^i$ are computed to make $\tilde{\underline{H}}_{k+\ell+1} Z^*$ and $\tilde{K}_{k+\ell+1}^i Z^*$ respectively upper-Hessenberg and upper-triangular again. At this point we can truncate (5) to obtain

$$\begin{aligned} AX_{k+1} &= U_{k+2} \underline{H}_{k+1} \\ L^i X_{k+1} &= V_{k+1}^i K_{k+1}^i \quad (i = 1, 2, \dots, \ell), \end{aligned}$$

and truncate $Z \tilde{\mathbf{c}}_{k+\ell+1}$ to obtain \mathbf{c}_{k+1} so that $\tilde{X}_{k+\ell+1} \tilde{\mathbf{c}}_{k+\ell+1} = X_{k+1} \mathbf{c}_{k+1}$. The truncation is expected to keep important components, since the directions removed from $X_{k+\ell+1}$ are perpendicular to the current best approximation \mathbf{x}_{k+1} , and also to the previous best approximations $\mathbf{x}_k, \mathbf{x}_{k-1}, \dots, \mathbf{x}_1$. If the rotation and truncation are combined in one step, then the

computational cost of the method is $\mathcal{O}((\ell + 1)(n + m + p^1 + \cdots + p^\ell))$, which quickly becomes smaller than the (re)orthogonalization cost as k grows.

To illustrate our approach, let us consider a one-parameter Tikhonov example where $\ell = 1$. First we expand $X_1 = \mathbf{x}_1$ with vectors $A^*A\mathbf{x}_1$ and $L^*L\mathbf{x}_1$. Let $A\tilde{X}_{1+2} = \tilde{U}_{2+2}\tilde{H}_{1+2}$ and $L\tilde{X}_{1+2} = \tilde{V}_{1+2}\tilde{K}_{1+2}$, and use \tilde{H}_{1+2} and \tilde{K}_{1+2} to compute $\tilde{\mathbf{c}}_{1+2}$. We then compute a rotation matrix Z_2 so that $Z_2\tilde{\mathbf{c}}_{2:3} = \pm\|\tilde{\mathbf{c}}_{2:3}\|\mathbf{e}_1$, and let Z be defined as in (6). The matrices $\tilde{H}_{1+2}Z^*$ and $\tilde{K}_{1+2}Z^*$ are no longer have their original structure, hence, we need to compute orthonormal P and Q such that $P\tilde{H}_{1+2}Z^*$ is again upper-Hessenberg and $Q\tilde{K}_{1+2}Z^*$ is upper-triangular. Schematically we have

$$\begin{array}{c} \xrightarrow{\tilde{\mathbf{c}}_{1+2}^*} \begin{bmatrix} \times & \times & \times \end{bmatrix} \xrightarrow{(Z\tilde{\mathbf{c}}_{1+2})^*} \begin{bmatrix} \times & \times & 0 \end{bmatrix} \\ \xrightarrow{\tilde{H}_{1+2}} \begin{bmatrix} \times & \times & \times \\ \times & \times & \times \\ 0 & \times & \times \\ 0 & 0 & \times \end{bmatrix} \xrightarrow{\tilde{H}_{1+2}Z^*} \begin{bmatrix} \times & \times & \times \\ \times & \times & \times \\ 0 & \times & \times \\ 0 & \times & \times \end{bmatrix} \xrightarrow{P\tilde{H}_{1+2}Z^*} \begin{bmatrix} \times & \times & \times \\ \times & \times & \times \\ 0 & \times & \times \\ 0 & 0 & \times \end{bmatrix} \\ \xrightarrow{\tilde{K}_{1+2}} \begin{bmatrix} \times & \times & \times \\ 0 & \times & \times \\ 0 & 0 & \times \end{bmatrix} \xrightarrow{\tilde{K}_{1+2}Z^*} \begin{bmatrix} \times & \times & \times \\ 0 & \times & \times \\ 0 & \times & \times \end{bmatrix} \xrightarrow{Q\tilde{K}_{1+2}Z^*} \begin{bmatrix} \times & \times & \times \\ 0 & \times & \times \\ 0 & 0 & \times \end{bmatrix} \end{array}$$

accompanied by the decompositions

$$\begin{aligned} A(\tilde{X}_{1+2}Z^*) &= (\tilde{U}_{2+2}P^*)(P\tilde{H}_{1+2}Z^*) \\ L(\tilde{X}_{1+2}Z^*) &= (\tilde{V}_{1+2}Q^*)(Q\tilde{K}_{1+2}Z^*). \end{aligned}$$

At this point we truncate the subspaces by removing the last columns from $\tilde{X}_{1+2}Z^*$, $\tilde{U}_{2+2}P^*$, $P\tilde{H}_{1+2}Z^*$, $\tilde{V}_{1+2}Q^*$, and $Q\tilde{K}_{1+2}Z^*$, and the bottom rows of $P\tilde{H}_{1+2}Z^*$ and $Q\tilde{K}_{1+2}Z^*$, to obtain

$$\begin{aligned} AX_2 &= U_3H_2 \\ LX_2 &= V_2K_2. \end{aligned}$$

Below we summarize the steps of the new algorithm for solving problem (1). In our implementation we take care to use full reorthogonalization and avoid extending X_k , U_{k+1} , and V_k^i with numerically linearly dependent vectors. We omit these steps from the pseudocode for brevity. In addition, we initially expand the search space solely with $A^*\mathbf{u}_{k+1}$ until the discrepancy principle can be satisfied conform Proposition 1 in Sect. 3.

Algorithm 2 (Multidirectional Tikhonov regularization)

Input: Measurement matrix A , regularization operators. L^1, \dots, L^ℓ , and data \mathbf{b} .

Output: Approximate solution $\mathbf{x}_k \approx \mathbf{x}_*$.

1. Initialize $\beta = \|\mathbf{b}\|$, $U_1 = \mathbf{b}/\beta$, $X_0 = []$, $\mathbf{x}_0 = \mathbf{0}$, and $\mu_0 = \mathbf{0}$.
for $k = 0, 1, \dots$, do
2. Expand X_k with $A^*A\mathbf{x}_k$, $L^1L^1\mathbf{x}_k, \dots, L^\ell L^\ell\mathbf{x}_k$.
3. Update $A\tilde{X}_{k+\ell+1} = \tilde{U}_{k+\ell+2}\tilde{H}_{k+\ell+1}$ and $L^i\tilde{X}_{k+\ell+1} = \tilde{V}_{k+\ell+1}^i\tilde{K}_{k+\ell+1}^i$.
4. Select μ_k ; see Sect. 4 and Algorithm 3.
5. $\tilde{\mathbf{c}}_{k+\ell+1} = \arg\min \left\| \left[\tilde{H}_{k+\ell+1}; \sqrt{\mu_k^1}\tilde{K}_{k+\ell+1}^1; \dots; \sqrt{\mu_k^\ell}\tilde{K}_{k+\ell+1}^\ell \right] \mathbf{c} - \beta\mathbf{e}_1 \right\|$.
6. Compute P , Q , and Z (see text).
7. Truncate $A(\tilde{X}_{k+\ell+1}Z^*) = (\tilde{U}_{k+\ell+2}P^*)(P\tilde{H}_{k+\ell+1}Z^*)$ to $AX_{k+1} = U_{k+2}H_{k+1}$.
Truncate $L^i(\tilde{X}_{k+\ell+1}Z^*) = (\tilde{V}_{k+\ell+1}^iQ^{i*})(Q^i\tilde{K}_{k+\ell+1}^iZ^*)$ to $L^iX_{k+1} = V_{k+1}^iK_{k+1}^i$.

8. Truncate $Z\tilde{\mathbf{c}}_{k+\ell+1}$ to obtain \mathbf{c}_{k+1} and set $\mathbf{x}_{k+1} = X_{k+1}\mathbf{c}_{k+1}$.
9. **if** $\|\mathbf{x}_{k+1} - \mathbf{x}_k\|/\|\mathbf{x}_k\|$ is sufficiently small **then break**

We have completed our discussion of the expansion and truncation phase of our algorithm. In the following section we discuss the extraction phase for one-parameter Tikhonov regularization and discuss the multiparameter case in later sections.

3 Parameter Selection in Standard Tikhonov

In this section we investigate parameter selection for general form one-parameter Tikhonov, where $\ell = 1$, $\mu = \mu^1$, and $L = L^1$. Multiple methods exist in the one-parameter case to determine particular μ_k , including the discrepancy principle, the L-curve criterion and generalized cross validation; see, for example, Hansen [11, Ch. 7]. We focus on the discrepancy principle which states that μ_k must satisfy

$$\|A\mathbf{x}_k(\mu_k) - \mathbf{b}\| = \eta\varepsilon, \quad (7)$$

where $\|\mathbf{e}\| \leq \varepsilon$ and $\eta > 1$ is a user supplied constant independent of ε .

Define the residual vector $\mathbf{r}_k(\mu) = A\mathbf{x}_k(\mu) - \mathbf{b}$ and the function $\varphi(\mu) = \|\mathbf{r}_k(\mu)\|^2$. A nonnegative μ_k satisfies the discrepancy principle if $\varphi(\mu_k) = \eta^2\varepsilon^2$. It is known that root finding methods can find solutions, for example, Lampe et al. [17] compare four of them. We prefer bisection for its reliability and straightforward analysis and implementation. The performance difference is not an issue because root finding requires a fraction of the total computation time and is no bottleneck. A unique solution μ_k exists under mild conditions, see for instance [3]. Below we give a proof using our own notation.

Assume \underline{H}_k and K_k are full rank and let $P_k \Sigma_k Q_k^*$ be the singular value decomposition of $\underline{H}_k K_k^{-1}$. Let the singular values be denoted by

$$\sigma_{\max} = \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_k = \sigma_{\min} > 0. \quad (8)$$

Now we can express $\mathbf{c}_k(\mu)$ and φ as

$$\begin{aligned} \mathbf{c}_k(\mu) &= (\underline{H}_k^* \underline{H}_k + \mu K_k^* K_k)^{-1} \underline{H}_k^* \beta \mathbf{e}_1 \\ &= K_k^{-1} (K_k^{-*} \underline{H}_k^* \underline{H}_k K_k^{-1} + \mu I)^{-1} K_k^{-*} \underline{H}_k^* \beta \mathbf{e}_1 \\ &= K_k^{-1} Q_k (\Sigma_k^2 + \mu I)^{-1} \Sigma_k P_k^* \beta \mathbf{e}_1 \end{aligned}$$

and

$$\begin{aligned} \varphi(\mu) &= \|\beta \mathbf{e}_1 - \underline{H}_k \mathbf{c}_k(\mu)\|^2 \\ &= \beta^2 \|\mathbf{e}_1 - \underline{H}_k K_k^{-1} Q_k (\Sigma_k^2 + \mu I)^{-1} \Sigma_k P_k^* \mathbf{e}_1\|^2 \\ &= \beta^2 \|(I - P_k P_k^*) \mathbf{e}_1 + P_k P_k^* \mathbf{e}_1 - P_k \Sigma_k (\Sigma_k^2 + \mu I)^{-1} \Sigma_k P_k^* \mathbf{e}_1\|^2 \\ &= \beta^2 \|(I - P_k P_k^*) \mathbf{e}_1\|^2 + \beta^2 \|\mu (\Sigma_k^2 + \mu I)^{-1} P_k^* \mathbf{e}_1\|^2. \end{aligned}$$

Or alternatively,

$$\varphi(\mu) = \beta^2 \|(I - P_k P_k^*) \mathbf{e}_1\|^2 + \beta^2 \sum_{j=1}^k \left(\frac{\mu}{\sigma_j^2 + \mu} \right)^2 |P_k|_{1j}^2. \quad (9)$$

Observe that P_k is a basis for the range of H_k and $I - P_k P_k^*$ is the orthogonal projection onto the nullspace $\mathcal{N}(H_k^*)$ and is sometimes denoted by $\mathcal{P}_{\mathcal{N}(H_k^*)}$. Furthermore, it can be verified that $H_k \beta e_1 \neq 0$ if $A^* b \neq 0$, that is, $b \notin \mathcal{N}(A^*)$.

Proposition 1 *If $\beta^2 \|(I - P_k P_k^*)e_1\|^2 \leq \eta^2 \varepsilon^2 < \|b\|^2$, then there exists a unique $\mu_k \geq 0$ such that $\varphi(\mu_k) = \eta^2 \varepsilon^2$.*

Proof (See also [3] and references therein). From (9) it follows that φ is a rational function with poles $\mu = -\sigma_j^2$ for all $\sigma_j > 0$, therefore, φ is C^∞ on the interval $[0, \infty)$. Additionally, φ is a strictly increasing and bounded function on the same interval, since

$$\frac{d}{d\mu} \left(\frac{\mu}{\sigma_j^2 + \mu} \right)^2 = 2 \frac{\mu \sigma_j^2}{(\sigma_j^2 + \mu)^3} > 0, \quad \text{for all } \mu > 0$$

implies $\varphi'(\mu) > 0$ and

$$\varphi(0) = \beta^2 \|(I - P_k P_k^*)e_1\|^2 \quad \text{and} \quad \lim_{\mu \rightarrow \infty} \varphi(\mu) = \beta^2 = \|b\|^2.$$

Consequently, there exists a unique $\mu_k \in [0, \infty)$ such that $\varphi(\mu_k) = \eta^2 \varepsilon^2$. \square

Beyond nonnegativity, the proposition above provides little insight on the location of μ_k on the real axis, and we would like to have lower and upper bounds. We determine bounds in Proposition 2 and believe the results to be new. Both in practice and for the proof of the subsequent proposition, it is useful to remove nonessential parts of $\varphi(\mu)$ and instead work with the function

$$\tilde{\varphi}(\mu) = \frac{\varphi(\mu) - \varphi(0)}{\beta^2} = \sum_{j=1}^k \left(\frac{\mu}{\sigma_j^2 + \mu} \right)^2 |P_k|_{1j}^2,$$

and the quantity

$$\tilde{\varepsilon}^2 = \frac{\eta^2 \varepsilon^2 - \varphi(0)}{\beta^2}. \quad (10)$$

Then $0 \leq \tilde{\varphi}(\mu) \leq \rho$, where $\rho = \|P_k^* e_1\| \leq 1$, and $\eta^2 \varepsilon^2$ satisfies the bounds in Proposition 1 if and only if $0 \leq \tilde{\varepsilon} < \rho$, and $\varphi(\mu_k) = \eta^2 \varepsilon^2$ if and only if $\tilde{\varphi}(\mu_k) = \tilde{\varepsilon}^2$.

Proposition 2 *If $0 \leq \tilde{\varepsilon} < \rho$, and μ_k is such that $\tilde{\varphi}(\mu_k) = \tilde{\varepsilon}^2$, then*

$$\frac{\tilde{\varepsilon}}{\rho - \tilde{\varepsilon}} \sigma_{\min}^2 \leq \mu_k \leq \frac{\tilde{\varepsilon}}{\rho - \tilde{\varepsilon}} \sigma_{\max}^2, \quad (11)$$

where σ_{\min} and σ_{\max} are as in (8).

Proof The key of the proof observe that

$$\frac{\mu}{\sigma_{\max}^2 + \mu} \leq \frac{\mu}{\sigma_j^2 + \mu} \leq \frac{\mu}{\sigma_{\min}^2 + \mu}$$

for all $j = 1, \dots, k$. Combining this observation with the definition of $\tilde{\varphi}$ yields

$$\left(\frac{\mu_k}{\sigma_{\max}^2 + \mu_k} \right)^2 \sum_{j=1}^k |P_k|_{1j}^2 \leq \sum_{j=1}^k \left(\frac{\mu_k}{\sigma_j^2 + \mu_k} \right)^2 |P_k|_{1j}^2 \leq \left(\frac{\mu_k}{\sigma_{\min}^2 + \mu_k} \right)^2 \sum_{j=1}^k |P_k|_{1j}^2,$$

Since $\sum_{j=1}^k |P_k|_{1j}^2 = \|P_k^* \mathbf{e}_1\|^2 = \rho^2$ and $\tilde{\varphi}(\mu_k) = \tilde{\varepsilon}^2$, it follows that

$$\frac{\mu_k}{\sigma_{\max}^2 + \mu_k} \rho \leq \tilde{\varepsilon} \leq \frac{\mu_k}{\sigma_{\min}^2 + \mu_k} \rho.$$

Hence, if $\tilde{\varepsilon} = 0$, then $\mu_k = 0$ and we are done. Otherwise $\mu_k \neq 0$ and we can divide by ρ , take the reciprocals, and subtract 1 to arrive at

$$\frac{\sigma_{\max}^2}{\mu_k} \geq \frac{\rho}{\tilde{\varepsilon}} - 1 \geq \frac{\sigma_{\min}^2}{\mu_k}.$$

So that

$$\frac{\mu_k}{\sigma_{\max}^2} \leq \frac{\tilde{\varepsilon}}{\rho - \tilde{\varepsilon}} \leq \frac{\mu_k}{\sigma_{\min}^2},$$

and the proposition follows. \square

It is undesirable to work with the inverse of K_k when it becomes ill-conditioned. Instead it may be preferred to use the generalized singular value decomposition (GSVD)

$$\begin{aligned} \underline{H}_k &= P_k C_k Z_k^{-1} \\ K_k &= Q_k S_k Z_k^{-1}, \end{aligned}$$

where P_k and Q_k have orthogonal columns and Z_k is nonsingular. The matrices C_k and S_k are diagonal with entries $0 \leq c_1 \leq c_2 \leq \dots \leq c_k$ and respectively $s_1 \geq \dots \geq s_k \geq 0$, such that $c_i^2 + s_i^2 = 1$. The generalized singular values are given by c_i/s_i and are understood to be infinite when $s_i = 0$. If K_k is nonsingular, then the generalized singular values coincide with the singular values of $\underline{H}_k K_k^{-1}$. See Golub and Van Loan [8, Section 8.7.3] for more information.

Using a similar derivation as before, we can show that

$$\varphi(\mu) = \beta^2 \|(I - P_k P_k^*) \mathbf{e}_1\|^2 + \beta^2 \sum_{j=1}^k \left(\frac{\mu s_j^2}{c_j^2 + \mu s_j^2} \right)^2 |P_k|_{1j}^2$$

and that the new bounds are given by

$$\frac{\tilde{\varepsilon}}{\rho - \tilde{\varepsilon}} \left(\frac{c_1}{s_1} \right)^2 \leq \mu_k \leq \frac{\tilde{\varepsilon}}{\rho - \tilde{\varepsilon}} \left(\frac{c_k}{s_k} \right)^2.$$

Here μ_k is unbounded from above if $s_k = 0$, that is, if K_k becomes singular.

The bounds in this section can be readily computed and used to implement bisection and the secant method. We consider parameter selection for multiparameter regularization in the following section.

4 A Multiparameter Selection Strategy

Choosing satisfactory μ_k^i in multiparameter regularization is more difficult than the corresponding one-parameter problem. See for example [1, 2, 6, 14, 16, 20, 20]. In particular, there is no obvious multiparameter extension of the discrepancy principle. Nevertheless, methods based on the discrepancy principle exist and we will discuss three of them.

Brezinski et al. [2] had some success with operators splitting. Substituting $\mu_k^i = v_k^i \omega_k^i$ in (3) with nonnegative weights ω_k^i and $\sum_{i=1}^{\ell} \omega_k^i = 1$ leads to

$$\operatorname{argmin}_{\mathbf{c}} \sum_{i=1}^{\ell} \omega_k^i (\|\underline{H}_k \mathbf{c} - \beta \mathbf{e}_1\|^2 + v_k^i \|K_k^i \mathbf{c}\|^2).$$

This form of the minimization problem suggests the approximation of $X_k^* \mathbf{x}_*$ by a linear combination [2, Sect. 3] of $\mathbf{c}_k^i(v_k^i)$, where

$$\mathbf{c}_k^i(v) = \operatorname{argmin}_{\mathbf{c}} \|\underline{H}_k \mathbf{c} - \beta \mathbf{e}_1\|^2 + v \|K_k^i \mathbf{c}\|^2 \quad (i = 1, 2, \dots, \ell), \quad (12)$$

and v_k^i is such that $\|\underline{H}_k \mathbf{c}_k^i(v_k^i) - \beta \mathbf{e}_1\| = \eta \varepsilon$. Alternatively, Brezinski et al. [2] consider solving

$$\mathbf{c}_k = \operatorname{argmin}_{\mathbf{c}} \left\| [\underline{H}_k; \sqrt{v_k^1} K_k^1; \dots; \sqrt{v_k^{\ell}} K_k^{\ell}] \mathbf{c} - \beta \mathbf{e}_1 \right\|,$$

where v^i are fixed and obtained from (12). The latter approach provides better results in exchange for an additional QR decomposition. In either case, operator splitting is a straightforward approach, but does not necessarily satisfy the discrepancy principle exactly.

Lu and Pereverzyev [19] and later Fornasier et al. [5] rewrite the constrained minimization problem as a differential equation and approximate

$$F(\boldsymbol{\mu}) = \|\underline{H}_k \mathbf{c}_k(\boldsymbol{\mu}) - \beta \mathbf{e}_1\|^2 + \sum_{i=1}^{\ell} \mu^i \|K_k^i \mathbf{c}_k(\boldsymbol{\mu})\|^2$$

by a model function $m(\boldsymbol{\mu})$ which admits a straightforward solution to the constructed differential equation. However, it is unclear which $\boldsymbol{\mu}$ the method finds and its solution may depend on the initial guess. On the other hand, it is possible to keep all but one parameter fixed and compute a value for the free parameter such that the discrepancy principle is satisfied. This allows one to trace discrepancy hypersurfaces to some extent.

Gazzola and Novati [6] describe another interesting method. They start with a one-parameter problem and successively add parameters in a novel way, until each parameter of the full multiparameter problem has a value assigned. Especially in early iterations the discrepancy principle is not satisfied, but the parameters are updated in each iteration so that the norm of the residual is expected to approach $\eta \varepsilon$. Unfortunately, we observed some issues in our implementation. For example, the quality of the result depends on initial values, as well as the order in which the operators are added (that is, the indexing of the operators). The latter problem is solved by a recently published and improved version of the method [7], which was brought to our attention during the revision of this paper.

We propose a new method that satisfies the discrepancy principle exactly, does not depend on an initial guess, and is independent of the scaling or indexing of the operators. The method uses the operator splitting approach in combination with new weights. Let us omit all k subscripts for the remainder of this section, and suppose $\mu^i = \mu \omega^i$, where ω^i are nonnegative, but do not necessarily sum to one, and μ is such that the discrepancy principle is satisfied. Then (3) can be written as

$$\operatorname{argmin}_{\mathbf{c}} \|\underline{H} \mathbf{c} - \beta \mathbf{e}_1\|^2 + \mu \sum_{i=1}^{\ell} \omega^i \|K^i \mathbf{c}\|^2. \quad (13)$$

Since the goal of regularization is to reduce sensitivity of the solution to noise, we use the weights

$$\omega^i = \frac{\|\mathbf{c}^i(v^i)\|}{\|D\mathbf{c}^i(v^i)\|}, \quad (14)$$

which bias the regularization parameters in the direction of lower sensitivity with respect to changes in v^i . Here D denotes the (total) derivative with respect to regularization parameter(s), and \mathbf{c}^i and v^i are defined as before, consequently

$$D\mathbf{c}^i(v^i) = -(\underline{H}^* \underline{H} + v^i K^{i*} K^i)^{-1} K^{i*} K^i \mathbf{c}^i(v^i).$$

If for some indices $D\mathbf{c}^i(v^i) = \mathbf{0}$, then we take a $\mathbf{c}^i(v^i)$ as the solution, or replace $\|D\mathbf{c}^i(v^i)\|$ by a small positive constant. With this parameter choice, the solution does not depend on the indexing of the operators, nor, up to a constant, on the scaling of A , \mathbf{b} , or any of the L^i . The former is easy to see; for the latter, let α , γ , and λ^i be positive constants, and consider the scaled problem

$$\operatorname{argmin}_{\hat{\mathbf{x}}} \|\gamma \mathbf{b} - \alpha A \hat{\mathbf{x}}\|^2 + \mu \sum_{i=1}^{\ell} \hat{\omega}^i \|\lambda^i L^i \hat{\mathbf{x}}\|^2.$$

The noisy component of $\gamma \mathbf{b}$ is $\gamma \mathbf{e}$ and $\|\gamma \mathbf{e}\| \leq \gamma \varepsilon$, hence the new discrepancy bound becomes

$$\|\alpha A \hat{\mathbf{x}} - \gamma \mathbf{b}\| = \gamma \eta \varepsilon.$$

The bound is satisfied when $\hat{\omega}^i = \alpha^2 / (\lambda^i)^2 \omega^i$, since in this case

$$\hat{\mathbf{x}} = \left(\alpha^2 A^* A + \mu \sum_{i=1}^{\ell} \omega^i \frac{\alpha^2}{(\lambda^i)^2} (\lambda^i)^2 L^{i*} L^i \right)^{-1} \alpha A^* \gamma \mathbf{b} = \frac{\gamma}{\alpha} \mathbf{x}.$$

and

$$\min_{\hat{\mathbf{x}}} \|\gamma \mathbf{b} - \alpha A \hat{\mathbf{x}}\|^2 + \mu \sum_{i=1}^{\ell} \hat{\omega}^i \|\lambda^i L^i \hat{\mathbf{x}}\|^2 = \gamma^2 \left(\min_{\mathbf{x}} \|A \mathbf{x} - \mathbf{b}\|^2 + \mu \sum_{i=1}^{\ell} \omega^i \|L^i \mathbf{x}\|^2 \right).$$

It may be checked that the weights in (14) are indeed proportional to $\alpha^2 / (\lambda^i)^2$, that is

$$\omega^i = \frac{\|\mathbf{c}^i(v^i)\|}{\|D\mathbf{c}^i(v^i)\|} \sim \frac{\alpha^2}{(\lambda^i)^2}.$$

There are additional viable choices for ω^i , including two smoothed versions of the above:

$$\omega^i = \frac{\|\underline{H} \mathbf{c}^i(v^i)\|}{\|\underline{H} D\mathbf{c}^i(v^i)\|} \quad \text{and} \quad \omega^i = \frac{\|K^i \mathbf{c}^i(v^i)\|}{\|K^i D\mathbf{c}^i(v^i)\|},$$

which consider the sensitivity of $\mathbf{c}^i(v^i)$ in the range of \underline{H} and K^i respectively. We summarize the new parameter selection in Algorithm 3 below.

Algorithm 3 (Multiparameter selection)

Input: Projected matrices \underline{H} , K^1, \dots, K^ℓ , $\beta = \|\mathbf{b}\|$, noise estimate ε , uncertainty parameter η , and threshold τ .

Output: Regularization parameters μ^1, \dots, μ^ℓ .

1. Use (12) to compute \mathbf{c}^i and v^i .
if $\|D\mathbf{c}^i(v^i)\| \leq \tau \|\mathbf{c}^i(v^i)\|$ **for some** i **then**

2. Set $\omega^i = \tau^{-1}$; or set $\mu^i = v^i$ and $\mu^j = 0$ for $j \neq i$.
else
3. Let $\omega^i = \|\mathbf{c}^i(v^i)\| / \|D\mathbf{c}^i(v^i)\|$.
4. Compute μ in (13) such that the discrepancy principle is satisfied.
5. Set $\mu^i = \mu\omega^i$.

An interesting property of Algorithm 3 is that, under certain conditions, $\mathbf{c}(\mu(\tilde{\varepsilon}))$ converges to the unregularized least squares solution

$$\mathbf{c}(\mathbf{0}) = (\underline{H}^* \underline{H})^{-1} \underline{H}^* \beta \mathbf{e}_1 = \underline{H}^+ \beta \mathbf{e}_1,$$

as the $\tilde{\varepsilon}$ goes to zero. Here \underline{H}^+ denotes the Moore–Penrose pseudoinverse and $\mathbf{c}(\mathbf{0})$ is the minimum norm solution of the unregularized problem. The following proposition formalizes this observation.

Proposition 3 Assume that \underline{H} is full rank, $\underline{H}^* \beta \mathbf{e}_1 \neq \mathbf{0}$, and that K^i is nonsingular for $i = 1, \dots, \ell$. Let $\tilde{\varepsilon}$ and ρ be defined as in Sect. 3, let $\eta > 1$ be fixed, and suppose that $v^i(\tilde{\varepsilon})$ and

$$\mu(\tilde{\varepsilon}) = (\mu^1(\tilde{\varepsilon}), \dots, \mu^\ell(\tilde{\varepsilon})) = \mu(\tilde{\varepsilon})(\omega^1(v^1(\tilde{\varepsilon})), \dots, \omega^\ell(v^\ell(\tilde{\varepsilon})))$$

are computed according to Algorithm 3 for all $0 \leq \tilde{\varepsilon} < \rho$. Then

$$\lim_{\tilde{\varepsilon} \downarrow 0} \omega^i(v^i(\tilde{\varepsilon})) = \omega^i(0) \quad \text{and} \quad \lim_{\tilde{\varepsilon} \downarrow 0} \mathbf{c}(\mu(\tilde{\varepsilon})) = \mathbf{c}(\mathbf{0}).$$

Proof First note that $\underline{H}^* \beta \mathbf{e}_1 \neq \mathbf{0}$ implies that $\beta > 0$ and $\rho > 0$. Since \underline{H} is full rank, the maps

$$v \mapsto \mathbf{c}^i(v), \quad v \mapsto D\mathbf{c}^i(v), \quad \text{and} \quad \mu \mapsto \mathbf{c}(\mu)$$

are continuous for all $v \geq 0$ and $\mu \geq \mathbf{0}$, where the latter bound should be interpreted elementwise. Hence

$$\lim_{v \downarrow 0} \mathbf{c}^i(v) = \mathbf{c}^i(0), \quad \lim_{v \downarrow 0} D\mathbf{c}^i(v) = D\mathbf{c}^i(0), \quad \text{and} \quad \lim_{\mu \downarrow \mathbf{0}} \mathbf{c}(\mu) = \mathbf{c}(\mathbf{0}).$$

It remains to be shown that

$$\lim_{\tilde{\varepsilon} \downarrow 0} v^i(\tilde{\varepsilon}) = 0, \quad \|D\mathbf{c}^i(0)\| \neq 0, \quad \text{and} \quad \lim_{\tilde{\varepsilon} \downarrow 0} \mu(\tilde{\varepsilon}) = \mathbf{0}. \quad (15)$$

Let $\tilde{\varepsilon}$ be restricted to the interval $[0, \rho/2]$ and define $v_{\max}^i = \sigma_{\max}^2(\underline{H}(K^i)^{-1})$. By Proposition 2,

$$0 \leq v^i(\tilde{\varepsilon}) \leq \frac{\tilde{\varepsilon}}{\rho - \tilde{\varepsilon}} v_{\max}^i \leq v_{\max}^i,$$

which proves the first limit in (15). Furthermore, using the definitions of $\mathbf{c}^i(v^i(\tilde{\varepsilon}))$ and $D\mathbf{c}^i(v^i(\tilde{\varepsilon}))$ we find the bounds

$$\begin{aligned} 0 &< \rho\beta \frac{\sigma_{\min}(\underline{H})}{\|\underline{H}\|^2 + v_{\max}^i \|K^i\|^2} \leq \|\mathbf{c}^i(v^i(\tilde{\varepsilon}))\| \leq \rho\beta \|\underline{H}^+ \mathbf{e}_1\|, \\ 0 &< \rho\beta \frac{\sigma_{\min}(\underline{H})\sigma_{\min}^2(K^i)}{(\|\underline{H}\|^2 + v_{\max}^i \|K^i\|^2)^2} \leq \|D\mathbf{c}^i(v^i(\tilde{\varepsilon}))\| \leq \rho\beta \frac{\|K^i\|^2 \|\underline{H}^+ \mathbf{e}_1\|}{\sigma_{\min}^2(\underline{H})}, \end{aligned}$$

which show that the inequality in (15) is satisfied. Moreover, the bounds show there exist ω_{\min} and ω_{\max} such that

$$0 < \omega_{\min} \leq \omega^i(\tilde{\varepsilon}) \leq \omega_{\max} < \infty.$$

Now, let $\mathbf{K}(\tilde{\varepsilon})$ be the nonsingular matrix satisfying

$$\mathbf{K}(\tilde{\varepsilon})^* \mathbf{K}(\tilde{\varepsilon}) = \sum_{i=1}^{\ell} \omega^i(\tilde{\varepsilon}) K^{i*} K^i,$$

then it can be checked that

$$\|\underline{H} \mathbf{K}(\tilde{\varepsilon})^{-1}\|^2 \leq \frac{\|\underline{H}\|^2}{\min_i \omega_{\min} \sigma_{\min}^2(K^i)} < \infty.$$

Define the right hand side of the equation above as M , then by Proposition 2, each entry of $\boldsymbol{\mu}(\tilde{\varepsilon})$ is bounded from below by 0 and from above by

$$\frac{\tilde{\varepsilon}}{\rho - \tilde{\varepsilon}} M \omega_{\max},$$

which goes to 0 as $\tilde{\varepsilon} \downarrow 0$. Therefore, this proves second limit in (15). \square

Proposition 3 is related to [9, Thm 3.3.3], where it is shown that the solution of a standard form Tikhonov regularization problem converges to a minimum norm least squares solution when the discrepancy principle is used and the noise converges to zero.

In this section we have discussed a new parameter selection method. In the next section we will look at the effect of perturbations in the parameters on the obtained solutions.

5 Perturbation Analysis

The goal of regularization is to make reconstruction robust with respect to noise. By extension, a high sensitivity to the regularization parameters is undesirable. Consider a set of perturbed parameters $\boldsymbol{\mu}_k + \Delta\boldsymbol{\mu}$; if $\|\Delta\boldsymbol{\mu}\|$ is sufficiently small

$$\begin{aligned} \mathbf{c}_k(\boldsymbol{\mu}_k + \Delta\boldsymbol{\mu}) &= \mathbf{c}_k(\boldsymbol{\mu}_k) + D\mathbf{c}_k(\boldsymbol{\mu}_k)\Delta\boldsymbol{\mu} + \mathcal{O}(\|\Delta\boldsymbol{\mu}\|^2) \\ &= \mathbf{c}_k(\boldsymbol{\mu}_k) - M^{-1}\Delta M\mathbf{c}_k(\boldsymbol{\mu}_k) + \mathcal{O}(\|\Delta\boldsymbol{\mu}\|^2), \end{aligned}$$

where M and ΔM are defined as

$$M = \underline{H}_k^* \underline{H}_k + \sum_{i=1}^{\ell} \mu_k^i K_k^{i*} K_k^i, \quad \Delta M = \sum_{i=1}^{\ell} \Delta \mu_k^i K_k^{i*} K_k^i. \quad (16)$$

Therefore, one might choose $\boldsymbol{\mu}_k$ to minimize the sensitivity measure

$$\|D\mathbf{c}_k(\boldsymbol{\mu}_k)\Delta\boldsymbol{\mu}\| = \|M^{-1}\Delta M\mathbf{c}_k(\boldsymbol{\mu}_k)\|.$$

To see the connection with the previous section, suppose that $\boldsymbol{\mu}_k = v_k^i \mathbf{e}_i$ and $\Delta\boldsymbol{\mu} = \pm \|\Delta\boldsymbol{\mu}\| \mathbf{e}_i$, then

$$\|M^{-1}\Delta M\| \geq \frac{\|M^{-1}\Delta M\mathbf{c}_k(\boldsymbol{\mu}_k)\|}{\|\mathbf{c}_k(\boldsymbol{\mu}_k)\|} = \frac{\|D\mathbf{c}_k(\boldsymbol{\mu}_k)\Delta\boldsymbol{\mu}\|}{\|\mathbf{c}_k(\boldsymbol{\mu}_k)\|} = \frac{\|D\mathbf{c}_k^i(v_k^i)\| \|\Delta\boldsymbol{\mu}\|}{\|\mathbf{c}_k^i(v_k^i)\|} = \frac{\|\Delta\boldsymbol{\mu}\|}{\omega_k^i}$$

Thus, larger weights ω_k^i correspond to smaller lower bounds on $\|M^{-1}\Delta M\|$. Having small lower bounds is desirable, since we show in Propositions 4 and 5 that minimizing $\|M^{-1}\Delta M\|$ is equivalent to minimizing upper bounds on the forward and backward errors respectively.

Proposition 4 Given regularization parameters μ_k^i and perturbations $\mu_\star^i = \mu_k^i + \Delta\mu_k^i$, let $\mathbf{c}_k = \mathbf{c}_k(\mu_k)$, $\mathbf{c}_\star = \mathbf{c}_k(\mu_\star)$, $\mathbf{x}_k = X_k \mathbf{c}_k$, and $\mathbf{x}_\star = X_k \mathbf{c}_\star$. Assume \underline{H}_k and all K_k^i are of full rank and define matrices M and ΔM as in (16). If M and $M + \Delta M$ are nonsingular and the $\Delta\mu_k^i$ are sufficiently small so that $\|M^{-1}\Delta M\| < 1$, then

$$\frac{\|\mathbf{x}_k - \mathbf{x}_\star\|}{\|\mathbf{x}_k\|} \leq \frac{\|M^{-1}\Delta M\|}{1 - \|M^{-1}\Delta M\|}.$$

Proof Observe that $\mathbf{c}_k = M^{-1}\underline{H}_k^* \beta \mathbf{e}_1$ and $\mathbf{c}_\star = (M + \Delta M)^{-1}\underline{H}_k^* \beta \mathbf{e}_1$. With a little manipulation we obtain

$$\mathbf{c}_\star = (M + \Delta M)^{-1} M \mathbf{c}_k = (I + M^{-1}\Delta M)^{-1} \mathbf{c}_k = \sum_{j=0}^{\infty} (-M^{-1}\Delta M)^j \mathbf{c}_k.$$

It follows that

$$\frac{\|\mathbf{c}_k - \mathbf{c}_\star\|}{\|\mathbf{c}_k\|} = \frac{1}{\|\mathbf{c}_k\|} \left\| \sum_{j=1}^{\infty} (-M^{-1}\Delta M)^j \mathbf{c}_k \right\| \leq \sum_{j=1}^{\infty} \|M^{-1}\Delta M\|^j \leq \frac{\|M^{-1}\Delta M\|}{1 - \|M^{-1}\Delta M\|}.$$

Since X_k has orthonormal columns, the result of the proposition follows. \square

One may wonder if it is possible to pick a vector \mathbf{f} close to $\beta \mathbf{e}_1$ such that

$$\mathbf{c}_k = (M + \Delta M)^{-1} \underline{H}_k^* \mathbf{f}.$$

Or in other words, given perturbed regularization parameters, is there a perturbation of $\beta \mathbf{e}_1$ such that the optimal approximation to the exact solution is obtained? The following proposition provides a positive answer.

Proposition 5 Under the assumptions of Proposition 4, there exist vectors \mathbf{f} and \mathbf{g} such that $\mathbf{c}_k = (M + \Delta M)^{-1} \underline{H}_k^* \mathbf{f}$ and $\mathbf{c}_\star = M^{-1} \underline{H}_k^* \mathbf{g}$. Furthermore, \mathbf{f} and \mathbf{g} satisfy

$$\begin{aligned} \frac{\|\beta \mathbf{e}_1 - \mathbf{f}\|}{\|\beta \mathbf{e}_1\|} &\leq \kappa(\underline{H}_k) \frac{\|M^{-1}\Delta M\|}{1 - \|M^{-1}\Delta M\|}, \\ \frac{\|\beta \mathbf{e}_1 - \mathbf{g}\|}{\|\beta \mathbf{e}_1\|} &\leq \kappa(\underline{H}_k) \|M^{-1}\Delta M\| \end{aligned}$$

where $\kappa(\underline{H}_k)$ is the condition number of \underline{H}_k .

Proof The vector \mathbf{f} is easy to derive using the ansatz

$$(M + \Delta M)^{-1} \underline{H}_k^* \mathbf{f} = M^{-1} \underline{H}_k^* \beta \mathbf{e}_1.$$

Let $\underline{H}_k = QR$ denote the reduced QR-decomposition of \underline{H}_k , then

$$R^* Q^* \mathbf{f} = (M + \Delta M) M^{-1} \underline{H}_k^* \beta \mathbf{e}_1,$$

and

$$\mathbf{f} = QR^{-*} (M + \Delta M) M^{-1} \underline{H}_k^* \beta \mathbf{e}_1 + (I - QQ^*) \mathbf{v}$$

for arbitrary \mathbf{v} . Indeed, it is easy to verify that the above vector satisfies

$$\mathbf{c}_k = (M + \Delta M)^{-1} \underline{H}_k^* \mathbf{f}.$$

If we choose $v = \beta e_1$, then

$$f = QR^{-*} \Delta M M^{-1} R^* Q^* \beta e_1 + \beta e_1$$

so that

$$\frac{\|\beta e_1 - f\|}{\|\beta e_1\|} = \|QR^{-*} \Delta M M^{-1} R^* Q^* e_1\| \leq \|R^{-*}\| \|R^*\| \|\Delta M M^{-1}\|.$$

Here $\|R^{-*}\| \|R^*\|$ is the condition number $\kappa(\underline{H}_k)$ and $\|\Delta M M^{-1}\| = \|M^{-1} \Delta M\|$, since both M and ΔM are symmetric. This proves the first part of the proposition.

The second part is analogous. In particular, we use the *ansatz*

$$M^{-1} \underline{H}_k^* g = (M + \Delta M)^{-1} \underline{H}_k^* \beta e_1$$

and derive

$$g = R^{-*} Q M (M + \Delta M)^{-1} \underline{H}_k^* \beta e_1 + (I - Q Q^*) \beta e_1.$$

Again it is easy to verify that $c_* = M^{-1} \underline{H}_k^* g$. Observe that g can be rewritten as

$$g = R^{-*} Q ((I + \Delta M M^{-1})^{-1} - I) R^* Q^* \beta e_1 + \beta e_1$$

such that

$$\begin{aligned} \frac{\|\beta e_1 - f\|}{\|\beta e_1\|} &= \|R^{-*} ((I + \Delta M M^{-1})^{-1} - I) R^* Q^* e_1\| \\ &\leq \|R^{-*}\| \|R^*\| \|(I + \Delta M M^{-1})^{-1} - I\|. \end{aligned}$$

Since $\|\Delta M M^{-1}\| = \|M^{-1} \Delta M\| < 1$, it follows that

$$\|(I + \Delta M M^{-1})^{-1} - I\| \leq \sum_{j=1}^{\infty} \|\Delta M M^{-1}\|^j = \frac{\|M^{-1} \Delta M\|}{1 - \|M^{-1} \Delta M\|},$$

which concludes the proof. \square

We have discussed forward and backward error bounds which help motivate our parameter choice. Now that we have investigated each of the three phases of our method, we are ready to show numerical results.

6 Numerical Experiments

We benchmark our algorithm with problems from Regularization Tools by Hansen [10]. Each problem provides an ill-conditioned $n \times n$ matrix A , a solution vector x_* of length n and a corresponding measured vector b . We let $n = 1024$ and add a noise vector e to b . The entries of e are drawn independently from the standard normal distribution. The noise vector is then scaled such that $\varepsilon = \|e\|$ equals $0.01\|b\|$ or $0.05\|b\|$ for 1 and 5 % noise respectively. We use $\eta = 1.01$ for the discrepancy bound in (7). We test the algorithms with 1000 different noise vectors for every triplet A , x_* , and b and report the median results.

The algorithms terminate when the relative difference between two subsequent approximations is less than 0.01, when x_{k+1} is (numerically) linear dependent in X_k , when both U_{k+1} and none of the V_k^i can be expanded, or when a maximum number of iterations is reached. For Algorithm 2 we use a maximum of 20 iterations and for Algorithm 1 a maximum of

Table 1 One-parameter Tikhonov regularization results

Noise Problem	1 %				5 %			
	E_{od}	E_{md}	ρ_E	ρ_{mv}	E_{od}	E_{md}	ρ_E	ρ_{mv}
Baart	1.73e−01	1.11e−01	6.44e−01	1.93e+00	2.91e−01	2.71e−01	9.33e−01	1.53e+00
Deriv2-1	2.44e−01	2.44e−01	1.00e+00	1.00e+00	3.32e−01	3.32e−01	1.00e+00	7.78e−01
Deriv2-2	2.35e−01	2.35e−01	1.00e+00	8.33e−01	3.22e−01	3.22e−01	9.99e−01	7.78e−01
Deriv2-3	4.35e−02	4.35e−02	1.00e+00	9.17e−01	7.97e−02	7.64e−02	9.59e−01	1.17e+00
Foxgood	3.31e−02	3.30e−02	9.98e−01	6.67e−01	6.64e−02	6.63e−02	9.98e−01	6.67e−01
Gravity-1	3.85e−02	3.41e−02	8.84e−01	1.08e+00	7.39e−02	6.86e−02	9.28e−01	1.11e+00
Gravity-2	5.53e−02	5.26e−02	9.51e−01	1.10e+00	8.66e−02	8.39e−02	9.69e−01	1.11e+00
Gravity-3	1.03e−01	9.21e−02	8.98e−01	1.08e+00	1.14e−01	1.10e−01	9.69e−01	1.11e+00
Heat	9.26e−02	9.12e−02	9.85e−01	1.05e+00	2.02e−01	1.91e−01	9.45e−01	1.37e+00
Phillips	2.50e−02	2.50e−02	1.00e+00	1.00e+00	4.52e−02	4.52e−02	9.99e−01	1.00e+00

$(\ell + 1) \times 20$ iterations. For the sake of a fair comparison, the algorithms return the best obtained approximations and their iteration numbers.

For each test problem, the tables below list the relative error obtained with Algorithm 1, abbreviated by E_{od} , and Algorithm 2, abbreviated by E_{md} . OD and MD stand for one direction and multidirectional respectively. Also listed are the ratio ρ_E of E_{md} to E_{od} and the ratio ρ_{mv} of the number of matrix-vector products. That is,

$$\rho_E = \frac{E_{md}}{E_{od}} \quad \text{and} \quad \rho_{mv} = \frac{\# \text{MVs Algorithm 2}}{\# \text{MVs Algorithm 1}}$$

Only matrix-vector multiplications with A , A^* , L^i , and L^{i*} count towards the total number of MVs used by each algorithm. We note, however, that multiplications with L^i and L^{i*} are often less costly than multiplications with A and A^* .

Table 1 lists the results one-parameter Tikhonov regularization, where we used the following regularization operators. The first derivative operator L_1 with stencil $[1, -1]$ for Gravity-3, Heat-5, Heat, and Phillips. The second derivative operator L_2 with stencil $[1, -2, 1]$ for Deriv2-1, Deriv2-2, Foxgood, Gravity-1, and Gravity-2. The third derivative operator L_3 with stencil $[-1, 3, -3, 1]$ for Baart. The fifth derivative operator L_5 with stencil $[-1, 5, -10, 10, -5, 1]$ and Deriv2-3. The derivative operators L_d are of size $(n - d) \times n$.

The table shows that multidirectional subspace expansion can obtain small improvements in the relative error at the cost of a small number of extra matrix-vector products, especially for 1 % noise. We stress that in these cases, Algorithm 1 is allowed to perform additional MVs, but converges with a higher relative error. If there is no improvement in the relative error, we see that multidirectional subspace expansion can improve convergence, for example, for the Deriv2 problems as well as Foxgood.

Table 2 lists the results for multiparameter Tikhonov regularization. We used the following regularization operators for each problem: the derivative operator L_d as listed above, the identity operator I , and the orthogonal projection $(I - N_d N_d^*)$, where the columns of N_d are an orthonormal basis for the nullspace $\mathcal{N}(L_d)$.

Overall, we observe larger improvements in the relative error for multidirectional subspace expansion, but also a larger number MVs. We no longer see cases where multidirectional

Table 2 Multiparameter Tikhonov regularization results

Noise	1 %				5 %			
Problem	E_{od}	E_{md}	ρE	ρ_{mv}	E_{od}	E_{md}	ρE	ρ_{mv}
Baart	1.72e-01	5.39e-02	3.12e-01	2.60e+00	2.84e-01	2.59e-01	9.14e-01	2.60e+00
Deriv2-1	2.27e-01	5.82e-03	2.56e-02	1.81e+00	3.21e-01	2.91e-02	9.08e-02	2.20e+00
Deriv2-2	2.29e-01	2.03e-02	8.84e-02	1.55e+00	2.95e-01	4.91e-02	1.66e-01	1.72e+00
Deriv2-3	4.35e-02	4.32e-02	9.93e-01	1.00e+00	7.71e-02	7.71e-02	1.00e+00	1.00e+00
Foxgood	3.29e-02	1.10e-02	3.35e-01	1.35e+00	6.26e-02	5.44e-02	8.69e-01	1.35e+00
Gravity-1	3.69e-02	1.83e-02	4.96e-01	1.18e+00	7.24e-02	4.52e-02	6.25e-01	1.63e+00
Gravity-2	5.52e-02	3.97e-02	7.19e-01	2.04e+00	8.52e-02	6.96e-02	8.17e-01	2.26e+00
Gravity-3	1.02e-01	9.24e-02	9.07e-01	1.89e+00	1.14e-01	1.08e-01	9.54e-01	1.72e+00
Heat	8.79e-02	8.77e-02	9.98e-01	1.19e+00	1.97e-01	1.83e-01	9.30e-01	1.40e+00
Phillips	2.49e-02	2.47e-02	9.90e-01	1.21e+00	4.08e-02	4.01e-02	9.83e-01	1.40e+00

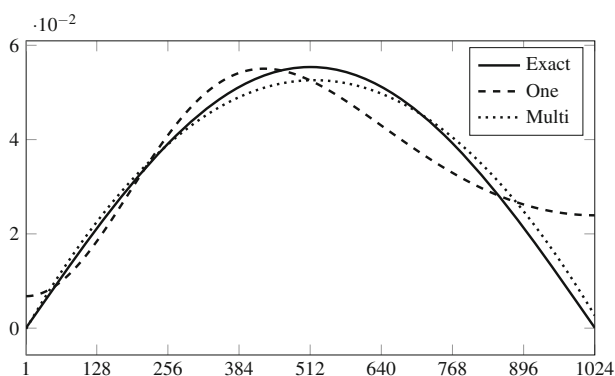


Fig. 1 baart test matrix with $n = 1024$ and 1% noise. The *solid line* is the exact solution. The *dashed line* is the solution obtained with multiparameter regularization and the residual subspace expansion (Algorithm 1). The *dotted line* is the solution obtained with multiparameter regularization and multidirectional subspace expansion (Algorithm 2)

subspace expansion terminates with fewer MVs. In fact, the relative error is the same for **Heat**, although more MVs are required. Finally, Fig. 1 illustrates an example of the improved results which can be obtained by using multidirectional subspace expansion.

In the next tests we attempt to reconstruct the original image from a blurred and noisy observation. Consider an $n \times n$ grayscale image with pixel values in the interval $[0, 1]$. Then \mathbf{x} is a vector of length n^2 obtained by stacking the columns of the image below each other. The matrix A represents a Gaussian blurring operator, generated with `blur` from Regularization Tools. The matrix A is block-Toeplitz with half-bandwidth `band=11` and the amount of blurring is given by the variance `sigma=5`. The entries of the noise vector \mathbf{e} are independently drawn from the standard normal distribution after which the vector is scaled such that $\varepsilon = \mathbb{E}[\|\mathbf{e}\|] = 0.05\|\mathbf{b}\|$. We take η such that $\|\mathbf{e}\| \leq \eta\varepsilon$ in 99.9% of the cases. That is,

$$\eta = 1 + \frac{3.090232}{\sqrt{2n^2}}.$$

For regularization we choose an approximation to the Perona–Malik [22] operator

$$\begin{aligned}\mathcal{L}(\mathbf{x}) &= \operatorname{div}(g(|\nabla \mathbf{x}|^2) \nabla \mathbf{x}), \\ g(s) &= e^{-s/\rho} \quad (\rho > 0),\end{aligned}$$

where ρ is a small positive constant. Because \mathcal{L} is a nonlinear operator, we first perform a small number of iterations with a finite difference approximation L_b of $\mathcal{L}(b)$. The resulting intermediate solution $\tilde{\mathbf{x}}$ is used for a new approximation $L_{\tilde{\mathbf{x}}}$ of $\mathcal{L}(\tilde{\mathbf{x}})$. Finally, we run the algorithms a second time with $L_{\tilde{\mathbf{x}}}$ and more iterations; see Reichel et al. [23] for more information regarding the implementation of the Perona–Malik operator.

The first test image is also used in [13, 23, 25], and is shown in Figure 2. We use $\rho = 0.075$, 20 iterations for the first run, and 100 iterations for the second run. The second image is an image of Saturn, see Figure 3. For this image we use $\rho = 0.03$, 25 iterations for the first run and 150 iterations for the second run. In both cases we stop the iterations around the point where convergence flattens out, as can be seen from the convergence history in Figure 4. The figure uses the peak signal-to-noise ratio (PSNR) given by

$$-20 \log_{10} \left(\frac{\|\mathbf{x}_* - \mathbf{x}_k\|}{n} \right)$$

versus the iteration number k . A higher PSNR means a higher quality reconstruction.

We observe that multidirectional subspace expansion may allow convergence to a more accurate solution. Because multidirectional subspace expansion requires extra matrix-vector products, we investigate the performance in Table 3 and when the PSNR of the output of Algorithm 2 achieves parity with the PSNR of the output of Algorithm 1. There is only a



Fig. 2 Deblurring results for Lizards. The original (left), observed (middle), and reconstructed images (right)



Fig. 3 Deblurring results for Saturn. The original (left), observed (middle), and reconstructed images (right)

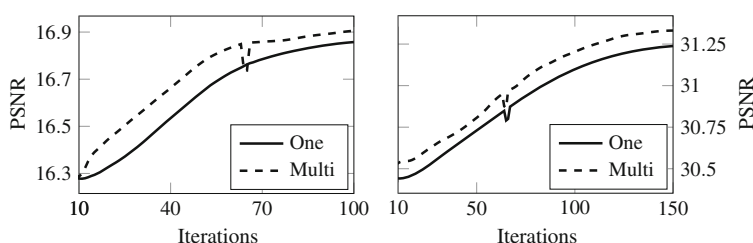


Fig. 4 Convergence history for Lizards (*left*) and Saturn (*right*)

Table 3 The number of matrix-vector products and wall clock time used by the different methods. The results in the upper rows are for Lizards and the results in the lower rows are for Saturn

Method	Total	A	A*	L	L*	Time (s)
Alg 1	399	100	100	100	99	30.9
Alg 2	581	191	100	191	99	38.7
Parity	395	129	69	129	68	23.5
Alg 1	599	150	150	150	149	82.3
Alg 2	889	295	150	295	149	98.4
Parity	637	211	108	211	107	62.3

small difference in the total number of matrix-vector products when parity is achieved, but a large improvement in wall clock time. This improvement is in large part due to the block operations which can only be used Algorithm 2. For reference, the runtimes were obtained on an Intel Core i7-3770 and with MATLAB R2015b on 64-bit Linux 4.2.5.

7 Conclusions

We have presented a new method for large-scale Tikhonov regularization problems. In accordance with Algorithm 2, the method combines a new multidirectional subspace expansion with optional truncation to produce a higher quality search space. The multidirectional expansion generates a richer search space, whereas the truncation ensures moderate growth. Numerical results illustrate that our method can yield more accurate results or faster convergence. Furthermore, we have presented lower and upper bounds on the regularization parameter when the discrepancy principle is applied to one-parameter regularization. These lower and upper bounds can be used in particular to initiate bisection or the secant method. In addition, we have introduced a straightforward parameter choice for multiparameter regularization, as summarized by Algorithm 3. The parameter selection satisfies the discrepancy principle, and is based on easy to compute derivatives that are related to the perturbation results of Sect. 5.

Acknowledgements We would like to thank the editor and the referees for their excellent feedback and helpful suggestions.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Belge, M., Kilmer, M.E., Miller, E.L.: Efficient determination of multiple regularization parameters in a generalized l-curve framework. *Inverse Probl.* **18**(4), 1161–1183 (2002)
2. Brezinski, C., Redivo-Zaglia, M., Rodriguez, G., Seatzu, S.: Multi-parameter regularization techniques for ill-conditioned linear systems. *Numer. Math.* **94**(2), 203–228 (2003)
3. Calvetti, D., Reichel, L.: Tikhonov regularization of large linear problems. *BIT* **43**(2), 263–283 (2003)
4. Fong, D., Saunders, M.A.: LSMR: an iterative algorithm for sparse least-squares problems. *SIAM J. Sci. Comput.* **33**(5), 2950–2971 (2011)
5. Fornasier, M., Naumova, V., Pereverzyev, S.V.: Parameter choice strategies for multipenalty regularization. *SIAM J. Numer. Anal.* **52**(4), 1770–1794 (2014)
6. Gazzola, S., Novati, P.: Multi-parameter Arnoldi–Tikhonov methods. *Electron. Trans. Numer. Anal.* **40**, 452–475 (2013)
7. Gazzola, S., Reichel, L.: A new framework for multi-parameter regularization. *BIT*, 1–31 (2015)
8. Golub, G.H., Van Loan, C.F.: *Matrix Computations*, 3rd edn. Johns Hopkins University Press, Baltimore (1996)
9. Groetsch, C.: *The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind*. Pitman Publishing, Boston (1984)
10. Hansen, P.C.: Regularization tools: a matlab package for analysis and solution of discrete ill-posed problems. *Numer. Algorithms* **6**, 1–35 (1994)
11. Hansen, P.C.: *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion*. SIAM, Philadelphia (1998)
12. Hochstenbach, M.E., Reichel, L.: An iterative method for Tikhonov regularization with a general linear regularization operator. *J. Integral Equ. Appl.* **22**(3), 465–482 (2010)
13. Hochstenbach, M.E., Reichel, L., Yu, X.: A Golub–Kahan-type reduction method for matrix pairs. *J. Sci. Comput.* **65**(2), 767–789 (2015)
14. Ito, K., Jin, B., Takeuchi, T.: Multi-parameter Tikhonov Regularization. [arXiv:1102.1173v2](https://arxiv.org/abs/1102.1173v2) [math.NA] (2011). Preprint
15. Kilmer, M.E., Hansen, P., Español, M.: A projection-based approach to general-form Tikhonov regularization. *SIAM J. Sci. Comput.* **29**(1), 315–330 (2007)
16. Kunisch, K., Pock, T.: A bilevel optimization approach for parameter learning in variational models. *SIAM J. Imaging Sci.* **6**(2), 938–983 (2013)
17. Lampe, J., Reichel, L., Voss, H.: Large-scale Tikhonov regularization via reduction by orthogonal projection. *Linear Algebra Appl.* **436**(8), 2845–2865 (2012)
18. Li, R.C., Ye, Q.: A Krylov subspace method for quadratic matrix polynomials with applications to constrained least squares problems. *SIAM J. Matrix Anal. Appl.* **25**(2), 405–528 (2003)
19. Lu, S., Pereverzyev, S.V.: Multi-parameter regularization and its numerical realization. *Numer. Math.* **118**(1), 1–31 (2011)
20. Lu, S., Pereverzyev, S.V., Shao, Y., Tautenhahn, U.: Discrepancy curves for multi-parameter regularization. *J. Inverse Ill-Posed Probl.* **18**(6), 655–676 (2010)
21. Paige, C.C., Saunders, M.A.: LSQR: an algorithm for sparse linear equations and sparse least squares. *ACM Trans. Math. Softw.* **8**(1), 43–71 (1982)
22. Perona, P., Malik, J.: Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Anal. Mach. Intell.* **12**(7), 629–639 (1990)
23. Reichel, L., Sgallari, F., Ye, Q.: Tikhonov regularization based on generalized Krylov subspace methods. *Appl. Numer. Math.* **62**(9), 1215–1228 (2012)
24. Reichel, L., Yu, X.: Matrix decompositions for Tikhonov regularization. *Electron. Trans. Numer. Anal.* **43**, 223–243 (2015)
25. Reichel, L., Yu, X.: Tikhonov regularization via flexible Arnoldi reduction. *BIT* **55**(4), 1145–1168 (2015)